Nonlinear Unmixing of Hyperspectral Images: Models and Algorithms

Nicolas Dobigeon, *Member, IEEE,* Jean-Yves Tourneret, *Senior Member, IEEE,* Cédric Richard, *Senior Member, IEEE,* José C. M. Bermudez, *Senior Member, IEEE,* Stephen McLaughlin, *Fellow, IEEE,* and Alfred O. Hero, *Fellow, IEEE*

Abstract

When considering the problem of unmixing hyperspectral images, most of the literature in the geoscience and image processing areas rely on the widely acknowledged linear mixing model (LMM). However, in specific but common contexts, the LMM may be not valid and other nonlinear models should be invoked. Consequently, over the last few years, several significant contributions have been proposed to overcome the limitations inherent in the LMM. In this paper, we present an overview of recent advances that deal with the nonlinear unmixing problem. The main nonlinear models are introduced and their validity discussed. Then, we describe the main classes of unmixing strategies designed to solve the problem in supervised and unsupervised frameworks. Finally, the problem of detecting nonlinear mixtures in hyperspectral images is addressed.

I. Introduction

Spectral unmixing (SU) is of prime interest for analyzing hyperspectral data, not only for remote sensing applications but also for planetary science and micro-spectroscopy. SU provides a comprehensive and quantitative mapping of the elementary materials that are present in the acquired data. More precisely, SU consists of identifying the spectral signatures of these materials (usually called

Part of this work has been funded by the Hypanema ANR Project n°ANR-12-BS03-003.

N. Dobigeon and J.-Y. Tourneret are with University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France. (e-mail: {nicolas.dobigeon, jean-yves.tourneret}@enseeiht.fr).

C. Richard is with Université de Nice Sophia-Antipolis, Observatoire de la Côte d'Azur, Laboratoire Lagrange, Parc Valrose, 06102 Nice Cedex 2, France. (e-mail: cedric.richard@unice.fr).

J. C. M. Bermudez is with the Department of Electrical Engineering, Federal University of Santa Catarina, Florianópolis, SC, Brazil (e-mail: j.bermudez@ieee.org).

S. McLaughlin is with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. (e-mail: s.mclaughlin@hw.ac.uk).

A. O. Hero is with University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA. (e-mail: hero@umich.edu).

endmembers) and estimating their relative contributions (or *abundances*) to the measured spectra. Due to the ill-posed nature, this is a challenging issue that has received considerable attention from the remote sensing, signal and image processing communities [1]. Hyperspectral data analysis can be supervised, when the endmembers are known, or unsupervised, when they are unknown. Irrespective of the case, most of the existing spectral mixture analysis approaches require the definition of the mixing model underlying the sensed observations. A mixing model describes in an analytical fashion how the endmembers combine to form the mixed spectrum measured by the sensor. The abundances parametrize the model. Given the mixing model, SU boils down to inverse this formation process to infer the quantities of interest, namely the endmembers and/or the abundances, from the collected spectra. Unfortunately, defining the direct observation model that links these meaningful quantities to the measured data is not a trivial issue, and requires a thorough comprehension of complex physical phenomena. A model based on radiative transfer (RT) could accurately describe the light scattering by the materials in the observed scene, but would probably lead to very complex unmixing problems. Fortunately, invoking simplifying but sound assumptions can lead to exploitable mixing models.

When the mixing scale is macroscopic and each photon reaching the sensor has interacted with just one material, the measured spectrum $\mathbf{y}_p \in \mathbb{R}^L$ in the pth pixel can be accurately described by the following linear mixing model (LMM),

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p}\mathbf{m}_r + \mathbf{n}_p \tag{1}$$

where L is the number of spectral bands, R is the number of endmembers present in the image, $\mathbf{m}_r$ is the spectral signatures of the rth endmember, $a_{r,p}$ is the abundance of the rth material in the pth pixel and $\mathbf{n}_p$ stands for an additive term associated with the measurement noise and the modeling error. Also, the abundances can be interpreted as the relative areas occupied by the materials in a given image pixel [2]. Thus it is natural to consider additional constraints regarding coefficients $a_{r,p}$

$$\begin{cases} a_{r,p} \geq 0, & \forall p,\ \forall r \\ \sum_{r=1}^{R} a_{r,p} = 1, & \forall p \end{cases} \tag{2}$$

In that case, SU can be formulated as a constrained blind source separation problem, or constrained linear regression, depending on the prior knowledge available regarding the endmember spectra.

Due to the relative simplicity of the model and the straightforward interpretation of the analysis results, LMM-based unmixing strategies predominate in the literature dedicated to the SU problem. The first automated spectral unmixing algorithms, proposed in the 1990's and based on geometrical concepts, were designed to identify endmembers as purest pixels. The most popular algorithms are probably the pixel purity index, N-FINDR and vertex component analysis (VCA) (see [1] for a comprehensive review of linear unmixing methods). It is worth noting that this class of algorithms does not explicitly rely on the assumption of pixels coming from linear mixtures. They only search

for endmembers as extremal points in the hyperspectral dataset, and this might indicate that some of these approaches can be still valid for nonlinear mixtures that preserve this property, provided there are pure pixels in the analyzed image.

Conversely, another class of geometrical algorithms focuses on minimizing the volume of the simplex spanned by the endmembers. Typical examples of these algorithms are the minimum volume transform (MVT), minimum volume simplex analysis (MVSA) or the simplex identification via variable splitting and augmented Lagrangian (SISAL) [1]. In that case, instances of endmembers as pure pixels are not required and the linearity assumption is implicitly invoked since the observations are assumed to lie inside the recovered simplex.

A large family of unmixing methods is based on a generative model, explicitly describing the observed pixels as linear mixtures of unknown endmembers. Joint estimation of the endmembers and the abundances can be formulated as a nonnegative matrix factorization (NMF) problem, whose additional constraints can be handled in deterministic or statistical framework. Among these fully parametric methods, one can point to the iterative constrained endmembers (ICE) algorithm, the minimum volume constrained NMF (MVC-NMF) or Bayesian algorithms.

All these techniques have been shown to be very useful whenever the LMM represents a good approximation to the actual mixing. There are, however, practical situations in which the LMM is not a suitable approximation [1]. In these cases, more complex mixing models need to be considered.

Several approximations to the RT theory have been proposed, such as Hapke's bidirectional model [2]. Unfortunately, these models require highly non-linear and integral formulations that hinder practical implementations of unmixing techniques. To circumvent these difficulties, several approximate but exploitable nonlinear mixing models have been recently proposed, and they have been receiving growing attention in the literature. This is due to their ability to capture important nonlinear effects that are inherent characteristics of hyperspectral images in several applications. They have proven essential to unveil meaningful information for the geoscience community. Developing effective unmixing algorithms based on nonlinear mixing models represents a challenge for the signal and image processing community. Supervised and unsupervised algorithms need to be designed to cope with nonlinear transformations that can be partially or totally unknown. Solving the nonlinear unmixing problem requires innovative approaches to existing signal processing techniques.

This article proposes to review the state-of-the-art nonlinear unmixing techniques. The paper is organized as follows. The existing nonlinear mixing models will be presented in Section II, where their validity and intrinsic limitations are discussed. The principal nonlinear unmixing algorithms are presented in Section III. Model-based and model-free algorithms are considered, and existing solutions for supervised and unsupervised unmixing are discussed. Section IV presents some recent strategies to detect nonlinear mixtures in hyperspectral data. Finally, challenges and future directions

for hyperspectral unmixing are reported in the concluding section.

II. Nonlinear models

In [1], it is explained that linear mixtures are reasonable when two assumptions are wholly fulfilled. First the mixing process must occur at a macroscopic scale [3]. Secondly, the photons that reach the sensor must interact with only one material, as is the case in checkerboard type scenes [4]. An illustration of this model is depicted in Fig. 1 for a scene composed of two materials. When one of these two assumptions does not hold, different nonlinear effects may occur. Two families of nonlinear models are described in what follows.

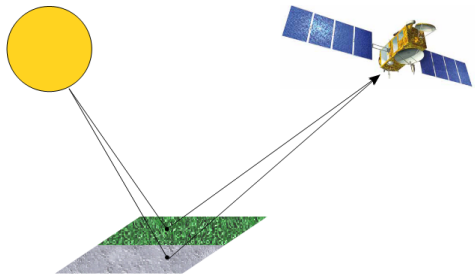

Fig. 1. Linear mixing model: the imaged pixel is composed of two materials.

A. Intimate mixtures

The first assumption for linear mixtures is a macroscopic mixing scale. However, there are common situations when interactions occur at a microscopic level. The spatial scales involved are typically smaller than the path length followed by the photons. The materials are said to be intimately mixed [2]. Such mixtures have been observed and studied for some time, e.g., for imaged scenes composed of sand or mineral mixtures [5]. They have been advocated for analyzing mixtures observed in the laboratory [6]. Based on RT theory, several theoretical frameworks have been derived to accurately describe the interactions suffered by the light when encountering surface composed of particles.

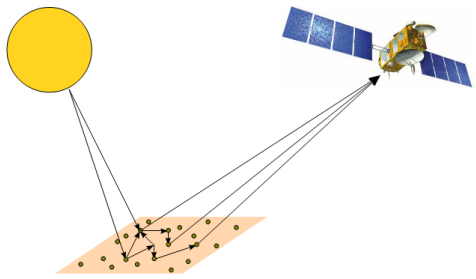

Fig. 2. Intimate mixture: the imaged pixel is composed of a microscopic mixture of several constituents.

An illustration of these interactions is represented in Fig. 2. Probably the most popular are the approaches developed by Hapke in [2] since they involve meaningful and interpretable quantities which have physical significance. Based on these concepts, several simplified nonlinear mixing models have been proposed to relate the measurement to some physical characteristics of the endmembers and to their corresponding abundances, which, for intimate mixtures, are associated with the relative mass fractions. In [6], the authors derive an analytical model to express the measured reflectances as a function of parameters intrinsic to the mixtures, e.g., the mass fraction, the characteristics of the individual particles (density, size) and the single-scattering albedo[1]. However this model also strongly depends on parameters inherent to the experiment since it requires the perfect knowledge of the geometric positioning of the sensor with respect to the observed sample. This dependency upon external parameters makes the inversion, i.e., the estimation of the mass fractions from the collected spectra, very difficult to implement.

More generally, it is worth noting that this first requirement of linear mixtures is intrinsically related to the definition of the endmembers, and this definition may be application dependent. Indeed, defining a pure material requires specification of the spatial or spectral resolution, which is application dependent. Consider a simple scene composed of 3 materials A, B and C. It is natural to expect retrieval of these components individually when analyzing the scene. However, in other circumstances, one may be interested in the material components themselves, for instance, A_1, A_2, B_1, B_2, C_1 and C_2 if we assume that each material is composed of 2 constituents. In that case, each pair of subcomponents combine and, by performing unmixing, one might also be interested in recovering each of these 6 components. Conversely, it may be well known that the material A can never be present in the observed scene without the material B. In such case, unmixing would consist of identifying the couple $A + B$ and C, without distinguishing the subcomponent A from the subcomponent B. This issue is frequently encountered in automated spectral unmixing. In each scenario, it is clear that more details are desired, the more the mixtures may not occur at a macroscopic scale. To circumvent this difficulty in defining the mixture scale, it makes sense to associate pure components with individual instances whose resolution has the same order of magnitude as the sensor resolution. For example, a patch of sand of spatially homogeneous composition can be considered as a unique pure component. In that case, most of the interactions occurring in most of the scenes of interest can be reasonably assumed to occur at a macroscopic level, at least when analyzing airborne and spaceborne remotely sensed images.

[1]Hapke's model and related approximations are not reproduced here for conciseness. Interested readers are invited to consult [6] or the more signal processing-oriented papers [7], [8] to have an understandable description of these models.

B. Bilinear models

Another type of nonlinear interaction is one that occurs at a macroscopic scale, in particular in so-called *multilayered* configurations. One may encounter this nonlinear model when the light scattered by a given material reflects off other materials before reaching the sensor. This is often the case for scenes acquired over forested areas, where there may be many interactions between the ground and the canopy. An archetypal example of this kind of scene is shown in Fig. 3.

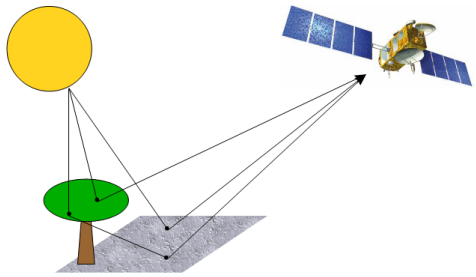

Fig. 3. Bilinear model: the imaged pixel is composed of two endmembers, namely tree and soil. In addition to the individual contribution of each material, bilinear interactions between the tree and the soil reach the sensor.

Several models have been proposed to analytically describe these interactions. They consist of including powers of products of reflectance. However they are usually employed such that interactions of orders greater than two are neglected. The resulting models are known as the family of the bilinear mixing models. Mathematically, for most of these bilinear models, the observed spectrum $\mathbf{y}_p \in \mathbb{R}^L$ in L spectral bands for the ith pixel is approximated by the following expansion

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p}\mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \beta_{i,j,p}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p. \tag{3}$$

where $\odot$ stands for the termwise (Hadamard) product:

$$\mathbf{m}_i \odot \mathbf{m}_j = \begin{pmatrix} m_{1,i} \\ \dots \\ m_{L,i} \end{pmatrix} \odot \begin{pmatrix} m_{1,j} \\ \dots \\ m_{L,j} \end{pmatrix} = \begin{pmatrix} m_{1,i}m_{1,j} \\ \dots \\ m_{L,i}m_{L,j} \end{pmatrix} \tag{4}$$

In the right-hand side of (3), the first term, also found in (1), summarizes the linear contribution in the mixing while the second term models nonlinear interactions between the materials. The coefficient $\beta_{i,j,p}$ adjusts the amount of nonlinearities between the components $\mathbf{m}_i$ and $\mathbf{m}_j$ in the pth pixel. Several alternatives for imposing constraints on these nonlinear coefficients have been suggested. In [9], Nascimento and Dias assume that the (linear) abundance and nonlinearity coefficients obey

$$\begin{cases} a_{r,p} \geq 0, & \forall p,\ \forall r \\ \beta_{i,j,p} \geq 0, & \forall p,\ \forall i \neq j \\ \sum_{r=1}^{R} a_{r,p} + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \beta_{i,j,p} = 1 \end{cases} \tag{5}$$

It is worth noting that, from (5), this Nascimento model (NM), also used in [10], can be interpreted as a linear mixing model with additional virtual endmembers. Indeed, considering $\mathbf{m}_i \odot \mathbf{m}_j$ as a pure component spectral signature with corresponding abundance $\beta_{i,j,p}$, the model in (5) can be rewritten

$$\mathbf{y}_p = \sum_{s=1}^{\tilde{R}} \tilde{a}_{s,p} \tilde{\mathbf{m}}_s + \mathbf{n}_p$$

with the positivity and additivity constraints in (2) where

$$\begin{cases} \tilde{a}_{s,p} \triangleq a_{r,p}, & \tilde{\mathbf{m}}_s \triangleq \mathbf{m}_r & s = 1, \ldots, R \\ \tilde{a}_{s,p} \triangleq \beta_{i,j,p}, & \tilde{\mathbf{m}}_s \triangleq \mathbf{m}_i \odot \mathbf{m}_j & s = R+1, \ldots, \tilde{R} \end{cases}$$

and $\tilde{R} = \frac{1}{2}R(R+1)$. This NM reduces to the LMM when $\tilde{a}_{s,p} = 0$ for $s = R+1, \ldots, \tilde{R}$.

Conversely, in [11], the authors have fixed the nonlinearity coefficients as functions of the (linear) abundance coefficients themselves: $\beta_{i,j,p} = a_{i,p}a_{j,p}$ $(i \neq j)$. The resulting model, called the Fan Model (FM) in what follows, is thus fully described by the mixing equation

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p}\mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} a_{i,p}a_{j,p}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p \tag{6}$$

subject to the constraints in (2). One argument to explain the direct relation between the abundances and the nonlinearity coefficients is the following: if the ith endmember is absent in the pth pixel, then $a_{i,p} = 0$ and there are no interactions between $\mathbf{m}_i$ and the other materials $\mathbf{m}_j$ $(j \neq i)$. More generally, it is quite natural to assume that the quantity of nonlinear interactions in a given pixel between two materials is directly related to the quantity of each material present in that pixel. However, it is clear that this model does not generalize the LMM, which can be a restrictive property.

More recently, to alleviate this issue, the generalized bilinear model (GBM) has been proposed in [12] by setting $\beta_{i,j,p} = \gamma_{i,j,p}a_{i,p}a_{j,p}$

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p}\mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{i,j,p}a_{i,p}a_{j,p}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p. \tag{7}$$

where the interaction coefficient $\gamma_{i,j,p} \in (0, 1)$ quantifies the nonlinear interaction between the spectral components $\mathbf{m}_i$ and $\mathbf{m}_j$. This model has the same interesting characteristic as the NM: the amount of nonlinear interactions is governed by the presence of the endmembers that linearly interact. In particular, again, if an endmember is absent in a pixel, there are no nonlinear interactions supporting this endmember. However, it also has the significant advantage of generalizing both the LMM when $\gamma_{i,j,p} = 0$ and the NM when $\gamma_{i,j,p} = 1$. Having $\gamma_{i,j,p} > 0$ indicate that only constructive interactions are considered.

For illustration, synthetic mixtures of $R = 3$ spectral components have been randomly generated according to the LMM, NM, FM and GBM. The resulting data set are represented in the space spanned by the three principal eigenvectors (associated with the three largest eigenvalues of the sample

covariance matrix of the data) identified by a principal component analysis in Fig. 4. These plots illustrate an interesting property: the spectral signatures of the pure components are still endmembers, i.e., vertices of the clusters, in the cases of FM and GBM mixtures contrary to the NM. In other words, endmember extraction algorithms based on geometrical interpretations, i.e., those that are looking for the simplex of largest or smallest volume (see [1] for details), can still be valid for the FM and the GBM.

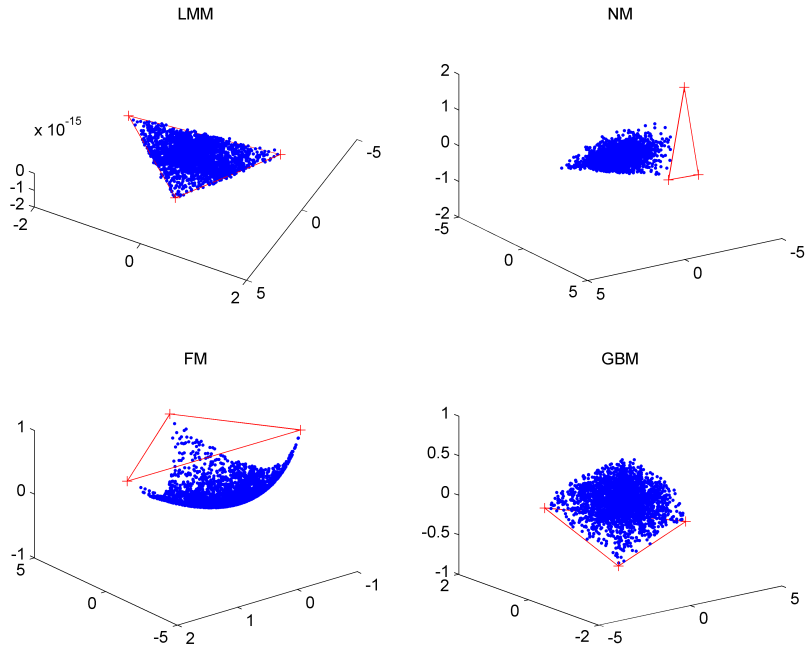

Fig. 4. Clusters of observations generated according to the LMM, the NM, the FM and the GBM (blue) and the corresponding endmembers (red).

All these bilinear models only include between-components interactions $\mathbf{m}_i \odot \mathbf{m}_j$ with $i \neq j$ but no within-components interactions $\mathbf{m}_i \odot \mathbf{m}_i$. However, in [13], the authors derive a nonlinear mixing model thanks to a thorough physical analysis of a simple canyon-like urban scene. Successive approximations and simplifying assumptions lead to the following linear-quadratic mixing model (LQMM)

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p}\mathbf{m}_r + \sum_{i=1}^{R}\sum_{j=i}^{R} \beta_{i,j,p}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p \tag{8}$$

with he positivity and additivity constraints in (2) and $\beta_{i,j,p} \in (0, 1)$. This model is similar to the FM in (6), with the noticeable difference that the nonlinear contribution includes quadratic terms $\mathbf{m}_i \odot \mathbf{m}_i$.

C. Other approximating physics-based models

To describe both macroscopic and microscopic mixtures, [14] introduces a dual model composed of two terms

$$\mathbf{y}_p = \sum_{r=1}^{R} a_{r,p}\mathbf{m}_r + a_{R+1,p}\mathcal{R}\left(\sum_{r=1}^{R} f_{r,p}\mathbf{w}_r\right) + \mathbf{n}_p.$$

The first term is similar to the one encountered in LMM and comes from the macroscopic mixing process. The second one, considered as an additional endmember with abundance $a_{R+1,p}$, describes the intimate mixture by the average single-scattering albedo expressed in the reflective domain by the mapping $\mathcal{R}(\cdot)$.

Altmann *et al.* have proposed in [15] an approximating physics-based model able to describe a wide class of nonlinearities. This model is derived by performing a second-order expansion using a post-nonlinear mixing model (PPNMM). More precisely, the pth observed pixel spectrum is defined as a nonlinear transformation $\mathbf{g}_p(\cdot)$ of a linear mixture of the endmember spectra

$$\mathbf{y}_p = \mathbf{g}_p\left(\sum_{r=1}^{R} a_{r,p}\mathbf{m}_r\right) + \mathbf{n}_p \tag{9}$$

where the nonlinear function $\mathbf{g}_p$ is defined as a second order polynomial nonlinearity parameterized by the unique nonlinearity parameter b_p

$$\begin{aligned} \mathbf{g}_p: \quad (0,1)^L &\to \mathbb{R}^L \\ \mathbf{x} &\mapsto \left[x_1 + b_p x_1^2, \ldots, x_L + b_p x_L^2\right]^T \end{aligned} \tag{10}$$

This model can be rewritten

$$\mathbf{y}_p = \mathbf{M}\mathbf{a}_p + b_p\left(\mathbf{M}\mathbf{a}_p\right) \odot \left(\mathbf{M}\mathbf{a}_p\right) + \mathbf{n}_p$$

where $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and $\mathbf{a}_p = [a_{1,p}, \ldots, a_{R,p}]^T$. The parameter b_p tunes the amount of nonlinearity present in the pth pixel of the image and this model reduces to the standard LMM when $b_p = 0$. Moreover, it can be easily shown that this polynomial post-nonlinear model (PPNM) includes bilinear terms $\mathbf{m}_i \odot \mathbf{m}_j$ $(i \neq j)$ similar to those defining the FM, NM and GBM, as well as quadratic terms $\mathbf{m}_i \odot \mathbf{m}_i$ similar to the LQMM in (8). This PNMM has been demonstrated to be sufficiently flexible to describe most of these bilinear models [15].

D. Discussion

Having reviewed the above physics-based models, an important remark must be made. It is important to note that these models do not take into account interactions from materials present in the neighborhood of the targeted pixel. This means that these bilinear models only consider scattering effects in a given pixel induced only by components that are present in this pixel. This is, naturally, a strong simplifying assumption that allows, at the end, the model parameters (abundance and nonlinear

coefficients) to be estimated pixel-by-pixel in the inversion step. Note however that the problem of taking into account adjacency effects, i.e., nonlinear interactions coming from spectral interference caused by atmospheric scattering, has been addressed in an unmixing context in [16].

III. Nonlinear unmixing algorithms

Significant and promising contributions have been proposed to nonlinearly unmix hyperspectral data. A wide class of nonlinear unmixing algorithms rely explicitly on a nonlinear physics-based parametric model, as detailed in Section II. Others do not require definition of the mixing model and rely on very mild assumptions regarding the nonlinearities. These two classes of algorithms, considered under the two unmixing scenarios (supervised vs. unsupervised) are described in this section.

A. Model-based parametric nonlinear unmixing algorithms

Given a nonlinear parametric model, SU can be formulated as a constrained nonlinear regression or a nonlinear source separation problem, depending on whether the endmember spectral signatures are known or not. When dealing with intimate mixtures, some authors propose converting the measured reflectance into a single scattering albedo average. Since this latter obeys a linear mixture, the mass fractions associated with each endmember can be estimated using a standard linear unmixing algorithm. This is the approach adopted in [6] and [7]. To avoid the painful conversion of the reflectance measurements, which would require the use of look-up tables, a common approach widely advocated consists of resorting to neural-networks (NN) to learn the nonlinear function. This is the strategy followed by Guilfoyle *et al.* in [17], for which several improvements have been proposed in [18] to reduce the computationally intensive learning step. Other NN-based algorithms include [19]–[22].

For the bilinear models introduced in Section II, supervised nonlinear optimization methods have been developed. When the observed pixel spectrum $\mathbf{y}_p$ is related to the parameters of interest $\boldsymbol{\theta}_p$ (a vector containing the abundance coefficients as well as any other nonlinearity parameters) through the function $\boldsymbol{\varphi}(\cdot)$, unmixing the pixel $\mathbf{y}_p$ consists of solving the following minimization problem

$$\hat{\boldsymbol{\theta}}_p = \underset{\theta}{\operatorname{argmin}} \|\mathbf{y}_p - \boldsymbol{\varphi}(\mathbf{M}; \boldsymbol{\theta})\|_2^2 \tag{11}$$

This problem raises two major issues: i) the nonlinearity of the criterion resulting from the underlying nonlinear model $\boldsymbol{\varphi}(\cdot)$ and ii) the constraints the parameters in $\boldsymbol{\theta}$ are subjected to. Since the NM can be interpreted as a linear mixing model with additional virtual endmembers, estimation of the parameters are conducted with a linear optimization method in [9]. In [11], [23] dedicated to FM and GBM, the authors propose to linearize the objective criterion via a first-order Taylor series expansion of $\boldsymbol{\varphi}(\cdot)$.

Then, the fully constrained least square (FCLS) algorithm in [24] can be used to estimate the vector of parameters $\boldsymbol{\theta}$. An alternative algorithmic scheme proposed in [23] consists of resorting to a gradient descent method where the step-size parameter is adjusted by a constrained line search procedure to ensure the constraints inherent to the mixing model. An alternative initially introduced in [12] for GBM is based on Monte Carlo approximations, developed in a fully Bayesian statistical framework. The Bayesian setting has the great advantage of providing a convenient way to include the parameter constraints within the estimation problem, by defining appropriate priors for the parameters. This strategy has been also implemented to conduct PNMM-based unmixing in [15].

When the spectral signatures involved in these bilinear models need also to be identified with the abundances and nonlinear coefficients jointly, more ambitious unmixing algorithms need to be designed. In [25], the authors differentiate the NM to implement updating rules that generalizes the SPICE algorithm introduced in [26] to solve the linear mixing model. Conversely, NMF-based iterative algorithms have been advocated in [27] for the GBM, and in [13] for the LQMM described in (8). Adopting a more geometrical point-of-view, Heylen and Scheunders propose in [28] an integral formulation to compute distances on the geodesic induced by the GBM. The underlying idea is to derive an endmember extraction algorithm that identifies the simplex of maximum volume inscribed on the manifold defined by the pixels.

B. Model-free nonlinear unmixing algorithms

When the nonlinearity is unknown, the problem becomes even more challenging. In such cases, a possible approach is to use a geometrical-based unmixing technique that relies on graph-based approximate geodesic distances [29], or manifold learning techniques [30], [31]. Methods based on support vector machine are also detailed in [32], [33]. Another promising approach is to resort to nonparametric methods based on Gaussian processes [34], or on reproducing kernels [35]–[40], to approximate the unknown nonlinearity. These two later techniques are described in what follows.

Nonlinear algorithms operating in reproducing kernel Hilbert spaces (RKHS) have been a topic of considerable interest in the machine learning community, and have proved their worth in solving nonlinear problems. Kernel-based methods have been widely considered for detection and classification in hyperspectral images. Surprisingly, nonlinear unmixing approaches operating in RKHS have been investigated in a less in-depth way. The algorithms derived in [35], [36] were mainly obtained by replacing each inner product between endmember spectra, in the cost functions to be optimized, by a kernel function. This can be viewed as a nonlinear distortion map applied to the spectral signature of each material, independently of their interactions. This principle may be extremely efficient in solving detection and classification problems as a proper distortion can increase the detectability or separability of some patterns. It is however of little physical interest in solving the unmixing problem

because the nonlinear nature of mixing is not only governed by individual spectral distortions, but also by nonlinear interactions of the materials. In [37], a new kernel-based paradigm was proposed to take the nonlinear interactions of the endmembers into account. It consists of considering the problem

$$\hat{\psi}_p(\cdot) = \arg\min_{\psi \in \mathcal{H}} \sum_{\ell=1}^{L} [y_{\ell,p} - \psi(\mathbf{m}_{\lambda_\ell})]^2 + \mu \|\psi\|_{\mathcal{H}}^2 \tag{12}$$

where $\mathbf{m}_{\lambda_\ell}$ stands for the vector of the endmember signatures at the ℓ-th frequency band, namely, $\mathbf{m}_{\lambda_\ell} = [m_{\ell,1}, \ldots, m_{\ell,R}]^T$ with $\mathcal{H}$ a given functional space, and μ a positive parameter that controls the trade-off between regularity of the function $\psi(\cdot)$ and fitting. It is interesting to note that (12) is the functional counterpart to (11), where $\psi(\cdot)$ defines the nonlinear interactions between the endmembers. Clearly, this strategy may fail if the functional space $\mathcal{H}$ is not chosen appropriately. A successful strategy is to define $\mathcal{H}$ as an RKHS in order to exploit the so-called kernel trick. Let $\kappa(\cdot\,,\cdot)$ be the reproducing kernel of $\mathcal{H}$. The RKHS $\mathcal{H}$ must be carefully selected via its kernel in order to make it flexible enough to capture wide classes of nonlinear relationships, and to reliably interpret a variety of experimental measurements. In order to extract the mixing ratios of the endmembers, the authors in [37] focus their attention on partially linear models. More precisely, the function $\psi(\cdot)$ in problem (12) is defined by an LMM parameterized by the abundance vector $\mathbf{a}$, combined with a nonparametric term, that is,

$$\psi(\mathbf{m}_{\lambda_\ell}) = \mathbf{a}^\top \mathbf{m}_{\lambda_\ell} + \psi_{\text{nlin}}(\mathbf{m}_{\lambda_\ell}) \tag{13}$$

possibly subject to the constraints in (2), where ψ_{nlin} can be any real-valued function of an RKHS denoted by $\mathcal{H}_{\text{nlin}}$. This model generalizes the standard LMM, and mimics the PPNM when $\mathcal{H}_{\text{nlin}}$ is defined to be the space of polynomial functions of degree 2. Remember that the latter is induced by the polynomial kernel $\kappa(\mathbf{m}_{\lambda_\ell}, \mathbf{m}_{\lambda_{\ell'}}) = (\mathbf{m}_{\lambda_\ell}^\top \mathbf{m}_{\lambda_{\ell'}})^q$ of degree $q = 2$. More complex interaction mechanisms can be considered by simply changing $\kappa(\mathbf{m}_{\lambda_\ell}, \mathbf{m}_{\lambda_{\ell'}})$. By virtue of the reproducing kernel machinery, the problem can still be solved in the framework of (12).

A number of interesting variants of the problem (12) can be constructed to incorporate prior information. In [37] for instance, the authors propose to substitute the regularization term $\mu\|\psi\|_{\mathcal{H}}^2$ by

$$\mu \left(\frac{1}{u} \|\mathbf{a}\|^2 + \frac{1}{1-u} \|\psi_{\text{nlin}}\|_{\mathcal{H}}^2 \right) \tag{14}$$

with $0 < u < 1$. Problem (12) can be solved efficiently with respect to the partially linear model (13) and u because it is convex. This regularization allows the automatic adjustment of the balance between the LMM and the nonlinear mixing submodel $\psi_{\text{nlin}}(\cdot)$, which paves the road to nonlinear mixture detectors described hereafter. Regularization can also be used for integrating spatial information into the hyperspectral unmixing problem. In [38], the authors have proposed using ℓ_1-type regularizer of the form [41] to promote piecewise-constant transitions in the fractional abundances that parameterize

the partially linear model (13). The regularizer incorporated in problem (12) is expressed as

$$\eta \sum_{q \in \mathcal{N}(p)} \|\mathbf{a}_p - \mathbf{a}_q\|_1 \tag{15}$$

where $\| \|_1$ denotes the vector ℓ_1-norm, and $\mathcal{N}(q)$ the set of neighbors of the pixel p. The optimization problem (12) remains convex but not smooth with respect to (15). A split-Bregman iteration algorithm can be used to overcome this drawback.

Another strategy by Altmann *et al.* considers a kernel-based approach for nonlinear SU based on a nonlinear dimensionality reduction using a Gaussian process latent variable model (GPLVM) [34]. The main advantage of GPLVMs is their capacity to accurately model many different nonlinearities. GPLVMs construct a smooth mapping from the space of fractional abundances to the space of observed mixed pixels that preserves dissimilarities. This strategy is also considered in [40] by Nguyen *et al.*, who solve the so-called pre-image problem [42] studied by the machine learning community. In the spectral unmixing context, it means that pixels that are spectrally different have different latent variables and thus different abundance vectors. However, preserving local distances is also interesting: spectrally close pixels are expected to have similar abundance vectors and thus similar latent variables. Several approaches have been proposed to preserve similarities, including back-constraints and locally linear embedding. In [34], the authors have used a particular form of kernel which extends the generalized bilinear model in (7). The proposed algorithm is unsupervised in the sense that the endmembers contained in the image and the mixing model are not known. Only the number of endmembers is assumed to be known. As a consequence, the parameters to be estimated are the kernel parameters, the endmember spectra and the abundances for all image pixels.

IV. Detecting nonlinear mixtures

Consideration of nonlinear effects in hyperspectral images can provide more accurate results in terms of endmember and abundance identification. However, working with nonlinear models generally requires a higher computational complexity than approaches based on the LMM. Thus, unmixing linearly mixed pixels using nonlinear models should be avoided. Consequently, it is of interest to devise techniques to detect nonlinearities in the mixing process before applying any unmixing method. Linearly mixed pixels can then be unmixed using linear unmixing techniques, leaving the application of more involved nonlinear unmixing methods to situations where they are really necessary. This section describes approaches that have been recently proposed to detect nonlinear mixing in hyperspectral images.

A. Investigation using surrogate data methods

A possible approach to test for nonlinear mixing in hyperspectral images is to consider the components of the observation vector as data series and employ existing time series analysis techniques

for nonlinearity detection. Recently surrogate data methods were proposed as a complement to the investigation of nonlinearity in hyperspectral imagery [43]. The general procedure of surrogate data was proposed in the early 1990's [44]. One first assumes that the observed data to be tested come from a specific hypothetical process that characterizes a null hypothesis $\mathcal{H}_0$. An ensemble of surrogate data sets are then generated which satisfy $\mathcal{H}_0$ and share given properties of the observed data. A discriminating statistic is then computed for each surrogate data set and for the observed data. The hypothesis $\mathcal{H}_0$ is rejected if the value of the test statistics for the observed data is not consistent with those provided by the surrogate data for a specified significance level. The key to success in applying the surrogate data method are the proper generation of the surrogate data and specification of the discriminating statistics.

The surrogates in [43] were generated using a technique called improved amplitude adjusted Fourier transform (AAFT) [45], which tries to match the distributions and the power spectra iteratively and reportedly leads to less false rejections than the standard AAFT. Two different discriminating statistics were tested on the available data obtained from a 4-m AVIRIS image: the third order moment of the difference between adjacent samples and the spectral angle between these samples. The conclusions of this study indicate that the results of tests using surrogate data methods as applied in [43] may be incorporated with other supporting evidence to decide about the linearity of the mixing of endmember contributions in hyperspectral images.

B. Detection using a polynomial post-nonlinear mixing model (PPNMM)

One interesting approach for nonlinearity detection is to assume a parametric nonlinear mixing model that can model different nonlinearities that could affect the mixing of the endmember contributions to the pixel hyperspectral observations. A model that has been successfully applied to this end is the post-nonlinear mixing model (PPNMM) studied in [15], [46].

PPNMM assumes the post-nonlinear mixing described in (9) with the polynomial nonlinearity $\mathbf{g}_p$ defined in (10). Hence, the nonlinearity is characterized by the parameter b_p for each pixel in the scene. This parameter can be estimated in conjunction with abundance vector $\mathbf{a}_p$ and the noise variance σ_p^2. Defining $s^2(\mathbf{a}_p, b_p, \sigma^2)$ as the variance of the estimator $\hat{b}_p$ of b_p, and using the properties of the maximum likelihood estimator, it makes sense to approximate the distribution of $\hat{b}_p$ by the following Gaussian distribution

$$\hat{b}_p \sim \mathcal{N}\left(b_p, s^2(a_p, b_p, \sigma^2)\right).$$

The nonlinearity detection problem has the two hypotheses

$$\begin{cases} \mathcal{H}_0 & : \quad \mathbf{y}_p \text{ is distributed according to the LMM (1)} \\ \mathcal{H}_1 & : \quad \mathbf{y}_p \text{ is distributed according to the PPNMM (9)} \end{cases} \tag{16}$$

Hypothesis $\mathcal{H}_0$ is characterized by $b_p = 0$ whereas nonlinear models ($\mathcal{H}_1$) correspond to $b_p \neq 0$. Then, (16) can be rewritten as

$$\begin{cases} \mathcal{H}_0 & : \quad \hat{b}_p \sim \mathcal{N}(0, s_0^2) \\ \mathcal{H}_1 & : \quad \hat{b}_p \sim \mathcal{N}(b_p, s_1^2) \end{cases} \tag{17}$$

where $s_0^2 = s^2(\mathbf{a}_p, 0, \sigma^2)$ and $s_1^2 = s^2(\mathbf{a}_p, b_p, \sigma^2)$ with $b_p \neq 0$.

Detection can be performed using the generalized likelihood ratio test. The test strategy is given by

$$T = \frac{\hat{b}_p^2}{s_0^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \tag{18}$$

As shown in [46], the statistic T is normally distributed under the two hypotheses. Consequently, the threshold η can be explicitly related to the probability of false alarm (PFA) and the probability of detection (PD), i.e., the power of the test.

However, this detection strategy assumes the prior knowledge of the variances s_0^2 and s_1^2. If this is not the case, Altmann *et al.* show in [46] that the test strategy can be modified to

$$\hat{T} = \frac{\hat{b}_p^2}{\hat{s}_0^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta^*. \tag{19}$$

where the estimate $\hat{s}_0^2$ can be calculated as

$$\hat{s}_0^2 = \mathrm{CCRLB}(0; \hat{\mathbf{a}}_p, \hat{\sigma}^2) \tag{20}$$

where CCRLB is the constrained Cramér-Rao lower-bound [47] on estimates of the parameter vector $\boldsymbol{\theta} = [\mathbf{a}_p^T, b_p, \sigma^2]^T$ under $\mathcal{H}_0$, and $\hat{\mathbf{a}}_p$ (resp. $\hat{\sigma}^2$) is the MLE of $\mathbf{a}_p$ (resp. σ^2). The performance of the resulting test is illustrated in Fig. 5 which shows the pixels detected as nonlinear when generated according to various mixing models (LMM, FM, GBM and PNMM).

C. Robust model-free detection

The detector discussed in the previous section assumes a specific nonlinear mixing model under the alternative hypothesis. There are, however, situations where the actual mixing does not obey any available model. It is also possible there is insufficient information to opt for any existing nonlinearity model. In these cases, it is interesting to address the problem of determining whether an observed pixel is a linear function of endmembers or results from a generic nonlinear mixing.

One may consider the LMM (1) and the hyperplane $\mathcal{P}$ defined by

$$\mathcal{P} : \left\{ \boldsymbol{z}_p \middle| \boldsymbol{z}_p = \mathbf{M}\mathbf{a}_p, \sum_{r=1}^{R} a_{r,p} = 1 \right\}. \tag{21}$$

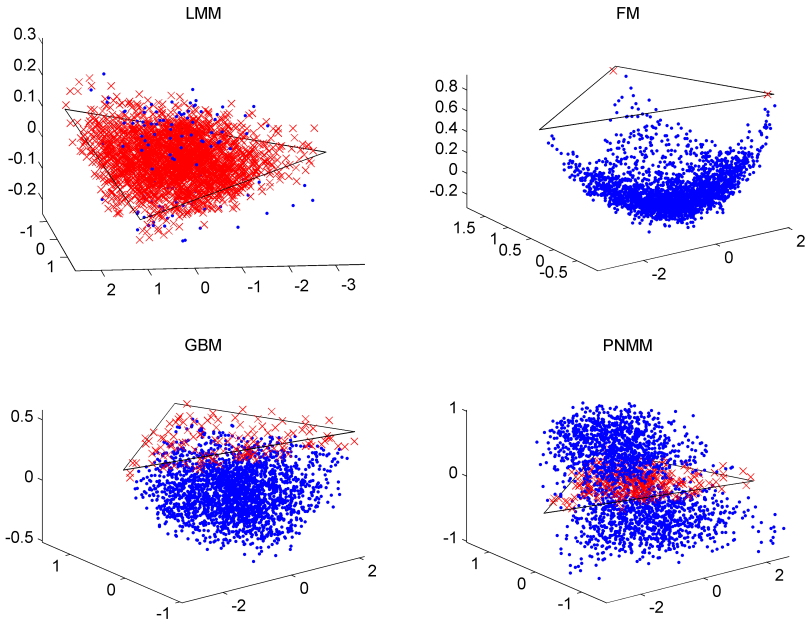

Fig. 5. Pixels detected as linear (red crosses) and nonlinear (blue dotted) for the four subimages generated according the LMM, FM, GBM, and PNMM. Black lines depict the simplex corresponding to the noise-free case LMM.

In the noise-free case, the hyperplane $\mathcal{P}$ lies in an $(R-1)$-dimensional subspace embedding all observations distributed according to the LMM. On the other hand, consider the general nonlinear mixing model[2]

$$\mathbf{y}_p = \mathbf{M}\mathbf{a}_p + \boldsymbol{\mu}_p + \mathbf{n}_p \tag{22}$$

where $\boldsymbol{\mu}_p$ is an $L \times 1$ deterministic vector that does not belong to $\mathcal{P}$, i.e., $\boldsymbol{\mu}_p \notin \mathcal{P}$ and $\mathbf{a}_p$ satisfies the constraints (2). Note that $\boldsymbol{\mu}_p$ can be a nonlinear function of the endmember matrix $\mathbf{M}$ and/or the abundance vector $\mathbf{a}_p$ and should be denoted as $\boldsymbol{\mu}_p(\mathbf{M}, \mathbf{a}_p)$ [49]. However, the arguments $\mathbf{M}$ and $\mathbf{a}_p$ are omitted here for brevity. Given an observation vector $\mathbf{y}_p$, the detection of nonlinear mixtures can be formulated as the following binary hypothesis testing problem

$$\begin{cases} \mathcal{H}_0 \;:\; \mathbf{y}_p \text{ is distributed according to the LMM (1)} \\ \mathcal{H}_1 \;:\; \mathbf{y}_p \text{ is distributed according to the model (22).} \end{cases}$$

Using the statistical properties of the noise $\mathbf{n}_p$ yields that $\mathrm{E}[\mathbf{y}_p|\mathcal{H}_0] = \mathbf{M}\mathbf{a}_p \in \mathcal{P}$ whereas $\mathrm{E}[\mathbf{y}_p|\mathcal{H}_1] = \mathbf{M}\mathbf{a}_p + \boldsymbol{\mu}_p \notin \mathcal{P}$. As a consequence, it makes sense to consider the squared Euclidean distance

$$\delta^2(\mathbf{y}_p) = \min_{\boldsymbol{z}\in\mathcal{P}} \|\mathbf{y}_p - \boldsymbol{z}\|^2 \tag{23}$$

[2]Note that a similar nonlinear mixing model coupled with a group-sparse constraint on $\boldsymbol{\mu}_p$ has been explicitly adopted in [48] to make more robust the unmixing of hyperspectral pixels when only a few of them are assumed to come from nonlinear mixtures.

between the observed pixel $\mathbf{y}_p$ and the hyperplane $\mathcal{P}$ to decide which hypothesis ($\mathcal{H}_0$ or $\mathcal{H}_1$) is true.

As shown in [49], the quantities $\delta^2(\mathbf{y}_p)$ are distributed according χ^2 distributions under the two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. The parameters of these distributions depend on the known matrix $\mathbf{M}$, the noise variance σ^2 and the nonlinearity vector $\boldsymbol{\mu}_p$. If σ^2 is known, the distribution of $\delta^2(\mathbf{y}_p)$ is perfectly known under $\mathcal{H}_0$ and partially known under $\mathcal{H}_1$. In this case, one may employ a statistical test that does not depend on $\boldsymbol{\mu}_p$, such as

$$T = \frac{\delta^2(\mathbf{y}_p)}{\sigma^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \tag{24}$$

As in the detection procedure detailed above IV-B, the threshold η can be related to the PFA and PD in closed-form expressions.

If the noise variance σ^2 is unknown, which is the case in most practical applications, one may replace σ^2 in (24) with an estimate $\hat{\sigma}^2$. This yields the test

$$T^* = \frac{\delta^2(\mathbf{y}_p)}{\hat{\sigma}^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta \tag{25}$$

where η is the threshold computed as previously. The PFA and PD of the test (25) are then explicitly obtained using cumulative distribution functions of the χ^2 distribution. It was shown in [49] the better the estimation of σ^2, the closer the distributions of T and T^* and thus the closer the performances of the tests (24) and (25). Several techniques can be used to estimate σ^2. For instance, $\hat{\sigma}^2$ has been estimated in [49] through an eigen-analysis of the sample covariance matrix of a set of pixels assumed to share the same variance. The value of $\hat{\sigma}^2$ was determined as the average of the smallest eigenvalues of the sample covariance matrix. The accuracy of the estimate is a function of the number of eigenvalues considered. It was shown in [49] that a PFA smaller (resp. larger) than P_{FA}^* is obtained if $\hat{\sigma}^2 > \sigma^2$ (resp. $\hat{\sigma}^2 < \sigma^2$).

V. Conclusion

To overcome the intrinsic limitations of the linear mixing model, several recent contributions have been made to the modeling of the physical processes that underly the hyperspectral observations. Some models are physically based, and they attempt to account for between-material interactions the photons may be subjected to before reaching the spectro-imager. Based on these models, several parametric algorithms have been proposed to solve the resulting nonlinear unmixing problem. Another class of unmixing algorithms attempts to avoid the use of any rigid nonlinear model by resorting to nonparametric machine learning-inspired techniques. The price to pay for handling nonlinear interactions induced by multiple scattering effects or intimate mixtures is the computational complexity and a possible degradation of unmixing performance when processing large hyperspectral images. To overcome these difficulties, one possible strategy consists of detecting pixels subjected to nonlinear mixtures in a pre-processing step. Detected linearly mixed pixels can then benefit from the huge

and reliable literature dedicated to the linear unmixing problem. The remaining pixels, detected as nonlinear, can then be the subject of particular attention. Finally, this paper demonstrates that the nonlinear unmixing problem represents an exciting issue that requires the expertise from signal and image processing researchers with various methodological backgrounds.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observations and Remote Sens.*, vol. 5, no. 2, pp. 354–379, April 2012.

[2] B. Hapke, *Theory of Reflectance and Emittance Spectroscopy*. Cambridge, UK: Cambridge Univ. Press, 1993.

[3] R. B. Singer and T. B. McCord, "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Proc. 10th Lunar and Planetary Sci. Conf.*, March 1979, pp. 1835–1848.

[4] R. N. Clark and T. L. Roush, "Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications," *J. Geophys. Res.*, vol. 89, no. 7, pp. 6329–6340, 1984.

[5] D. B. Nash and J. E. Conel, "Spectral reflectance systematics for mixtures of powdered hypersthene, labradorite, and ilmenite," *J. Geophys. Res.*, vol. 79, pp. 1615–1621, 1974.

[6] J. F. Mustard and C. M. Pieters, "Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra," *J. Geophys. Res.*, vol. 94, no. B10, pp. 13,619–13,634, Oct. 1989.

[7] J. M. P. Nascimento and J. M. Bioucas-Dias, "Unmixing hyperspectral intimate mixtures," in *Proc. SPIE Image and Signal Processing for Remote Sensing XVI*, L. Bruzzone, Ed., vol. 74830. SPIE, Oct. 2010, p. 78300C.

[8] R. Close, P. Gader, A. Zare, J. Wilson, and D. Dranishnikov, "Endmember extraction using the physics-based multi-mixture pixel model," in *Proc. SPIE Imaging Spectrometry XVII*, S. S. Shen and P. E. Lewis, Eds., vol. 8515. San Diego, California, USA: SPIE, Aug. 2012, pp. 85 150L–14.

[9] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE Image and Signal Processing for Remote Sensing XV*, L. Bruzzone, C. Notarnicola, and F. Posa, Eds., vol. 7477, no. 1. SPIE, 2009, p. 74770I.

[10] N. Raksuntorn and Q. Du, "Nonlinear spectral mixture analysis for hyperspectral imagery in an unknown environment," *IEEE Geosci. and Remote Sensing Lett.*, vol. 7, no. 4, pp. 836–840, 2010.

[11] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Int. J. Remote Sens.*, vol. 30, no. 11, pp. 2951–2962, June 2009.

[12] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. and Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[13] I. Meganem, P. Déliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear-quadratic mixing model for reflectances in urban environments," *IEEE Trans. Geosci. and Remote Sensing*, 2013, to appear.

[14] R. Close, P. Gader, J. Wilson, and A. Zare, "Using physics-based macroscopic and microscopic mixture models for hyperspectral pixel unmixing," in *Proc. SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, S. S. Shen and P. E. Lewis, Eds., vol. 8390. Baltimore, Maryland, USA: SPIE, April 2012, pp. 83 901L–83 901L–13.

[15] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, June 2012.

[16] D. Burazerovic, R. Heylen, B. Geens, S. Sterckx, and P. Scheunders, "Detecting the adjacency effect in hyperspectral imagery with spectral unmixing techniques," *IEEE J. Sel. Topics Appl. Earth Observations and Remote Sens.*, 2013, to appear.

[17] K. J. Guilfoyle, M. L. Althouse, and C.-I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Trans. Geosci. and Remote Sensing*, vol. 39, no. 8, pp. 2314–2318, Aug. 2001.

[18] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using radial basis functions and orthogonal least squares," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Vancouver, Canada, July 2011, pp. 1151–1154.

[19] J. Plaza, P. Martinez, R. Perez, and A. Plaza, "Nonlinear neural network mixture models for fractional abundance estimation in AVIRIS hyperspectral images," in *Proc. XIII NASA/Jet Propulsion Laboratory Airborne Earth Science Workshop*, Pasadena, CA, USA, 2004.

[20] J. Plaza, A. Plaza, R. Perez, and P. Martinez, "Automated generation of semi-labeled training samples for nonlinear neural network-based abundance estimation in hyperspectral data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, 2005, pp. 345–350.

[21] J. Plaza, A. Plaza, R. Pérez, and P. Martinez, "Joint linear/nonlinear spectral unmixing of hyperspectral image data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, 2007, pp. 4037–4040.

[22] J. Plaza, A. Plaza, Rosa, Perez, and P. Martinez, "On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images," *Pattern Recognition*, vol. 42, pp. 3032–3045, 2009.

[23] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unmixing hyperspectral images using the generalized bilinear model," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Vancouver, Canada, July 2011, pp. 1886–1889.

[24] D. C. Heinz and C. -I Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 29, no. 3, pp. 529–545, March 2001.

[25] P. Gader, D. Dranishnikov, A. Zare, and J. Chanussot, "A sparsity promoting bilinear unmixing model," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Shanghai, China, June 2012.

[26] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geosci. and Remote Sensing Lett.*, vol. 4, no. 3, pp. 446–450, July 2007.

[27] N. Yokoya, J. Chanussot, and A. Iwasaki, "Generalized bilinear model based nonlinear unmixing using semi-nonnegative matrix factorization," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Munich, Germany, 2012, pp. 1365–1368.

[28] R. Heylen and P. Scheunders, "Calculation of geodesic distances in nonlinear mixing models: Application to the generalized bilinear model," *IEEE Geosci. and Remote Sensing Lett.*, vol. 9, no. 4, pp. 644–648, July 2012.

[29] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 534–542, June 2011.

[30] H. Nguyen, C. Richard, P. Honeine, and C. Theys, "Hyperspectral image unmixing using manifold learning methods. derivations and comparative tests," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Munich, Germany, July 2012.

[31] G. Licciardi, X. Ceamanos, S. Doute, and J. Chanussot, "Unsupervised nonlinear spectral unmixing by means of NLPCA applied to hyperspectral imagery," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Munich, Germany, July 2012, pp. 1369–1372.

[32] J. Plaza, A. Plaza, P. Martinez, and R. Perez, "Nonlinear mixture models for analyzing laboratory simulated-forest hyperspectral data," *Proc. SPIE Image and Signal Processing for Remote Sensing IX*, vol. 5238, pp. 480–487, 2004.

[33] P.-X. Li, B. WU, and L. Zhang, "Abundance estimation from hyperspectral image based on probabilistic outputs of multi-class support vector machines," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, 2005, pp. 4315–4318.

[34] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, 2013, to appear.

[35] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Barcelona, Spain, July 2007, pp. 4041–4044.

[36] J. Broadwater and A. Banerjee, "A comparison of kernel functions for intimate mixture models," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Aug. 2009, pp. 1–4.

[37] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 480–492, Jan. 2013.

[38] ——, "Nonlinear abundance estimation of hyperspectral images with L1-norm spatial regularization," *IEEE Trans. Geosci. and Remote Sensing*, 2013, submitted.

[39] X. Li, J. Cui, and L. Zhao, "Blind nonlinear hyperspectral unmixing based on constrained kernel nonnegative matrix factorization," *Signal, Image and Video Processing*, pp. 1–13, 2012.

[40] N. H. Nguyen, J. Chen, C. Richard, P. Honeine, and C. Theys, "Supervised nonlinear unmixing of hyperspectral images using a pre-image methods," *EAS Publications Series*, vol. 59, pp. 417–437, 2013.

[41] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[42] P. Honeine and C. Richard, "Preimage problem in kernel-based machine learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 77–88, March 2011.

[43] T. Han and D. G. Goodenough, "Investigation of nonlinearity in hyperspectral imagery using surrogate data methods," *IEEE Trans. Geosci. and Remote Sensing*, vol. 46, pp. 2840–2847, Oct. 2008.

[44] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D-Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 77–94, Sept. 1992.

[45] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Physical Review Letters*, vol. 77, no. 4, pp. 635–638, July 1996.

[46] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinearity detection in hyperspectral images using a polynomial post-nonlinear mixing model," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1267–1276, April 2013.

[47] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285 –1301, Nov. 1990.

[48] N. Dobigeon and C. Févotte, "Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Gainesville, FL, June 2013, submitted.

[49] Y. Altmann, N. Dobigeon, J.-Y. Tourneret, and J. C. M. Bermudez, "A robust test for nonlinear mixture detection in hyperspectral images," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, June 2013.